Markov Decision Processes

states : S set of states

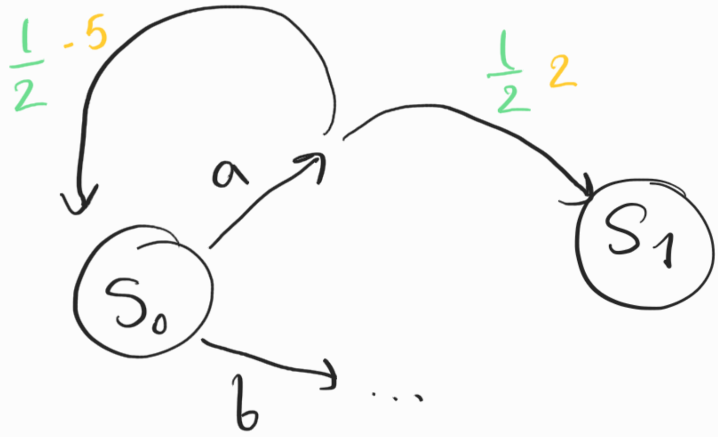
actions : A set of actions

transition function : $\Delta : S \times A \to \mathcal{D}ist(S \times \mathbb{R})$

rewards

$\Delta(s, a)(s', r)$: probability that from state s playing action a, we go to state s' and get reward r

probability
reward



$\Delta(s', r \mid s, a)$

Strategy ≡ policy :

$\pi : S \to A$ deterministic

or $\pi : S \to \mathrm{Dist}(A)$ stochastic

distributions

play ≡ trajectory ≡ path :

$\rho = S(0), A(0), R(0), S(1), A(1), R(1), \ldots$

return of a trajectory

$$G = \sum_{t \geq 0} \gamma^t R(t) = R(0) + \gamma R(1) + \gamma^2 R(2) + \ldots$$

$\gamma \in (0,1)$

Two Cases:

- either eventually we reach a sink

$$G = \sum_{t=0}^{\infty} R(t) \quad \text{is actually finite}$$

$\rightarrow$ FINITE HORIZON

- or the trajectory may be infinite

$$G = \sum_{t=0}^{\infty} \gamma^t R(t)$$

$\rightarrow$ DISCOUNTED

$\gamma \in (0,1)$: fixed constant

$$G = \sum_{t=0}^{\infty} \gamma^t R(t) = R(0) + \gamma R(1) + \gamma^2 R(2) + \gamma^3 R(3) \dots$$

$$\gamma^t \xrightarrow[t \to \infty]{} 0$$

Goal: Construct a strategy π

maximising $\mathbb{E}\left[G \mid S(0) = s_0 \wedge \pi \right]$

$s_0 \in S$ initial state